

## 修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 電気通信学研究科 情報通信工学専攻 博士前期課程		
氏 名	岸本 貴之	学籍番号	0730020
論 文 題 目	CRF による係り受け解析の結果を反映させた日本語形態素解析		
<p>要 旨</p> <p>形態素解析は日本語文章理解や情報検索において基礎となる技術である。近年、日本語形態素解析において、隠れマルコフモデル (HMM) などの従来法に代わる新しい手法として、Conditional Random Fields (CRF) による解析が提案され、最高水準の結果を示している。しかし、このモデルを用いても、誤った形態素を含んだ文がまだまだ多く存在しているのが現状である。日本語の文章を理解する上では、一つの文で正しい形態素列を得ているかどうかは重要であり、この意味で一文全てに誤りのない割合(文正解率)が一つの指標となる。</p> <p>本稿では、文正解率を上げることを目指して、新たに CRF による係り受け解析の結果を反映した CRF による形態素解析手法を提案する。従来の形態素解析では、隣接する 2 個までの単語間の相関から解析する。しかし、実際にはもっと広い範囲での単語の相関に影響されており、隣接する 2 単語だけでは解析できない事例が存在する。一方、係り受け解析では、文中に含まれる全ての文節間から解析を行うため、係り受け解析の結果を反映させることで、形態素解析の弱点を補完することが可能となる。文節の係り受け構造を決定するためには、正確な形態素を与える必要があるため、形態素の選択が適したものであるほど、係り受け解析の確かさもより大きくなると考えられる。このため、CRF による形態素解析で得られた尤度の高いいくつかの候補に対して、CRF による係り受け解析が最も良くなるような形態素系列を選べば、形態素解析の精度を上げることが出来ると期待される。</p> <p>提案モデルと従来モデルとの比較実験の結果、実用に十分な時間のロスで精度改善を行うことができた。形態素解析の確からしさの上位 10 件について、係り受け解析の結果を反映させた場合では、全単語情報、品詞名、単語区切りの文正解率がそれぞれ 1.97%, 2.38%, 1.39%向上させることができた。</p> <p>キーワード：形態素解析, 係り受け解析, Conditional Random Fields</p>			